Dis 12 Hypothesis testing

Thursday, 15 April 2021 9:00 AM

Agenda

1. Intro to binary hypothesis testing (Neyman Pearson formulation)
   - Example
   - Likelihood ratio test
   - Neyman Pearson lemma
2. MAP hypothesis test
   - Guided derivation

Obs. Y come up w/ "good" decision rule $r(Y)$ to decide btw 2 hypotheses H_0, H_1
null hypothesis — alt.

What is good?

$r: \mathbb{R} \to \{0,1\}$

2 Types of errors ← decision rule

Type 1 error : False Positive $r(y)=1 \mid X=0$

Type 2 error : False negative $r(y)=0 \mid X=1$

'Good' decision rule : (under NP)
minimizes Prob. of type II error with a constraint on Prob. of type I
upper bound on $\Pr(\text{type I}) \le \beta$

Ex: good doctor

⇒ decision rule r on whether a patient has cancer or doesn't have ← H_1

cancer ← H_0

obs: symptom of patient (Y)

$r(y) \to \begin{cases} 0 \to H_0 \\ 1 \to H_1 \end{cases}$

type I: patient actually doesn't have cancer, $r(y)=1$

type 2: $X=1$, $r(y)=0$

NP lemma

Q: What decision rule is optimal?

① Likelihood ratio test

$$L(y) = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)}$$

$$r(y) = \begin{cases} 1 & \text{if } L(y) > \lambda \\ 0 & \text{if } L(y) < \lambda \\ \text{Bern}(\gamma) & \text{if } L(y) = \lambda \end{cases}$$

What is λ? γ?

optimal decision rule

NP Lemma: opt. r — to find λ & γ, choose λ & γ s.t.

$$\Pr(r(Y)=1 \mid X=0) = \beta$$

$\Pr(\text{type I error}) = \beta$
'false +ve' — equal to bound

1. Hypothesis Testing for Bernoulli Random Variables

Assume that

- If $X = 0$, then $Y \sim \text{Bernoulli}(1/4)$. $\Pr(Y=1 \mid X=0)$
- If $X = 1$, then $Y \sim \text{Bernoulli}(3/4)$.

Using the Neyman-Pearson formulation of hypothesis testing, find the optimal randomized *decision rule* $r : \{0,1\} \to \{0,1\}$ with respect to the criterion

$$\min_{\text{randomized } r:\{0,1\}\to\{0,1\}} \Pr(r(Y) = 0 \mid X = 1)$$
$$\text{s.t. } \Pr(r(Y) = 1 \mid X = 0) \le \beta,$$

where $\beta \in [0,1]$ is a given upper bound on the false positive probability.

y is either 0 or 1 — $P(Y=y)$

$$L(y) = \frac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)} = \frac{(\frac{3}{4})^y (\frac{1}{4})^{1-y}}{(\frac{1}{4})^y (\frac{3}{4})^{1-y}} = \frac{3^y \cdot 1}{1 \cdot 3^{1-y}} = 3^y \cdot 3^{-(1-y)} = 3^{-1+2y}$$

by definition

Sanity check: $Y \sim \text{Bern}(\frac{3}{4})$, $\Pr(Y=1) = \frac{3}{4}$

$$L(y) = 3^{2y-1} = \begin{cases} 3 & \text{if } y=1 \\ \frac{1}{3} & \text{if } y=0 \end{cases}$$

$L(0)$

opt. decision rule

$$L(Y) = 3^{2Y-1} = \begin{cases} 3 & \text{if } Y=1 \\ \frac{1}{3} & \text{if } Y=0 \end{cases}$$

$\Leftrightarrow r(Y) = \begin{cases} 1 & \text{if } L(Y) > \lambda \\ 0 & \vdots \\ \text{Bern}(\gamma) & \vdots \end{cases}$

$L(y) = 3 = \begin{cases} 3 & \text{if } y=1 \\ \frac{1}{3} & \text{if } y=0 \end{cases}$ ← $L(0)$

opt. decision rule

$\Longleftrightarrow r(Y) = \begin{cases} 1 & \text{if } L(Y) > \lambda \\ 0 & \vdots \\ Bern(\gamma) & \end{cases}$

$r(y) = \begin{cases} 1 & \text{if } L(y) > \lambda \\ 0 & \text{if } L(y) < \lambda \\ Bern(\gamma) & \text{if } L(y) = \lambda \end{cases}$ $\equiv Pr(\text{type I error}) = \beta$

where $\boxed{Pr(\boxed{r(Y) = 1} \mid X = 0) = \beta}$ (equals to upper bound)

$\quad ||| \quad (L(Y) > \lambda) \cup (L(Y) = \lambda \cap Bern(\gamma) = 1)$

$Pr(L(Y) > \lambda) + Pr(L(Y) = \lambda \cap Bern(\gamma) = 1 \mid X = 0)$

$\beta \in [0,1]$

if $\lambda = \frac{1}{2}$ $\quad \phi$

$= Pr(L(Y) > \frac{1}{2} \mid X=0) + Pr(L(Y) = \frac{1}{2} \cap Bern(\gamma) = 1 \mid X = 0)$, where the second term $= 0$

$= Pr(L(Y) > \frac{1}{2} \mid X=0) = Pr(L(Y) = 3 \mid X = 0) = Pr(Y=1 \mid X=0)$

if $\beta = \frac{1}{4}$ then $\lambda = \frac{1}{2}$ results in opt decision rule

if $\beta = \frac{1}{4}$, we're looking for λ, γ s.t.

$P(\text{type I error}) = \frac{1}{4}$

if $\lambda \in (\frac{1}{3}, 3)$

if $\lambda = 3$

Want $P(t1\ err) = \beta$, check what β's we can 'hit' by tuning γ

$\phi = \cancel{Pr(L(Y) > 3)} + Pr(L(Y) = 3 \cap Bern(\gamma) = 1 \mid X = 0)$, where the struck term $= 0$

$= Pr(Bern(\gamma) = 1 \mid X = 0, L(Y = 3)) \times Pr(L(Y) = 3 \mid X = 0)$

$= (\gamma) \times Pr(Y = 1 \mid X = 0)$ $\quad Bern(\frac{1}{4})$

$= \gamma \times \frac{1}{4}$ $\quad \gamma \in [0,1]$ ∵ Bern

$P(t1\ err) = \frac{\gamma}{4} = \beta$

if $\lambda = 3$, can cover β where $\beta \in [0, \frac{1}{4}]$

if $\beta \in (\frac{1}{4}, 1) \Rightarrow (\lambda = \frac{1}{3})$, γ you calculate

2. MAP Hypothesis Testing

Let X have prior probabilities $\Pr(X=0) = \pi_0, \Pr(X=1) = \pi_1$, and Y be the observed variable.

(a) Recall that the MAP (binary) hypothesis test says:

$$\hat{X}_{MAP}(y) = \arg\max_x p_{X|Y}(x|y) = \begin{cases} 1 & \text{if } \frac{p_{Y|X}(y|1)}{p_{Y|X}(y|0)} \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{else} \end{cases}$$

Show that the MAP test $\hat{X}_{MAP}$ minimizes the probability of error $P(\hat{X}(Y) \neq X)$ over all tests $\hat{X}$.

(b) Suppose instead of minimizing the expected probability of error, we want to minimize a weighted sum of the Type-I and Type-II error, i.e. minimize:

$$C_0 \Pr(\hat{X}(Y) = 1 | X = 0) + C_1 \Pr(\hat{X}(Y) = 0 | X = 1)$$

for some $C_0 \geq 0$ and $C_1 \geq 0$, what test should we use?

$Pr(t1\ err) =$

$\quad Pr(L(Y) > 3 \cup$

$\lambda = \frac{1}{3}$

$Pr(t1\ err) = Pr(L(Y) > 3 \cup L(Y) = 3 \cap Bern(\gamma) = 1 \mid X = 0)$ $\quad \phi$

$= Pr(L(Y) = 3) +$

3. Gaussians and the MSE

Suppose you draw n i.i.d. data points $(x_1, y_1), \ldots, (x_n, y_n)$, where n is a positive integer and the true relationship is $Y = WX + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. (That is, Y has a linear dependence on X, with additive Gaussian noise.) Show that finding the MLE estimate of W given the data points $\{(x_i, y_i) : i = 1, \ldots, n\}$ is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2$$